Pergamon

0042-6989(95)00250-2

Vision Res., Vol. 36, No. 7, pp. 1007–1022, 1996

Printed in Great Britain
0042-6989/96 $15.00 + .00

Reduced Perceptual Dimensionality in Extrafoveal Vision

MARTIN JÜTTNER,*† INGO RENTSCHLER*

Received 15 September 1994; in revised form 15 March 1995; in final form 5 July 1995

The classification behaviour of human observers with respect to compound Gabor signals is tested at foveal and extrafoveal retinal positions. Classification performance is analysed in terms of a probabilistic classification model recently proposed by Rentschler, Jüttner and Caelli [(1994) *Vision Research, 34*, 669–687]. The analysis allows inferences about structure and dimensionality of the individual internal representations underlying the classification task and their temporal evolution during the learning process. Using this technique it is found that the internal representations of direct and eccentric viewing are intrinsically incommensurable, in the sense that extrafoveal pattern representations are characterized by a lower perceptual dimension in feature space relative to the corresponding physical input signals, whereas foveal representations are not. The observed deficits cannot be renormalized by size scaling (cortical magnification); however, they can be partially reduced by learning although the learning progress strongly depends on the observer's practice. The structural incommensurability between foveal and extrafoveal representations poses constraints on possible forms of foveal–extrafoveal interaction, which might have implications on related perceptual phenomena such as visual stability across saccadic eye movements.

Peripheral vision Cortical magnification Classification Bayesian classifiers Supervised learning Internal representation Visual stability

INTRODUCTION

Since the early work of Aubert and Foerster (1857) and Wertheim (1894) it is well-known that under photopic conditions, visual performance of peripheral vision is degraded compared to foveal vision in most tasks. In psychophysics the deficits in indirect viewing have been mostly characterized in terms of threshold indices such as grating acuity (Wertheim, 1894), Snellen acuity (Ludvigh, 1941; Weymouth, 1958; Virsu, Näsänen & Osmoviita, 1987), grating contrast sensitivity (Hilz & Cavonius, 1974; Kelly, 1984; Koenderink, Bouman, Bueno de Mesquita & Slappendel, 1978; Virsu & Rovamo, 1979; Virsu, Rovamo, Laurinen & Näsänen, 1982), vernier acuity (Hering, 1899; Westheimer, 1982; Levi, Klein & Aitsebaomo, 1985; Weymouth, 1958) and spatial phase resolution (Harvey, Rentschler & Weiss, 1985). These acuity measures reveal a dependence on eccentricity, which resembles the radially anisotropic mapping from the retinal photoreceptors to striate cortex.

To describe the retino-cortical mapping, Daniel and Whitteridge (1961) introduced the notion of the cortical magnification factor, M_C, which is defined as millimetres of cortex per degree of visual angle. Anatomical studies show that the inverse of M_C decreases approximately as a linear function of retinal eccentricity (Cowey & Rolls, 1974; Daniel & Whitteridge, 1961; Van Essen, Newsome & Maunsell, 1984). The variation of M_C with eccentricity suggests that many visual functions might be equated in central and peripheral vision by applying a scale factor F being proportional to M_C^{-1} to each linear dimension of the peripheral stimulus. This idea was first proposed by Koenderink *et al.* (1978), and by Rovamo and Virsu (1979). These authors showed that such a linear transformation could account for the variation of the grating contrast sensitivity with eccentricity.

The concept of size scaling has proved to apply well to tasks which, as those listed above, involve the detection of simple stimuli (e.g. sine-wave gratings) or the measurement of some acuity index (for a review see Wilson, Levi, Maffei, Rovamo & DeValois, 1990). However, there is experimental evidence that for complex stimuli the recognition performance in peripheral view cannot always be renormalized by size scaling alone. Strasburger, Harvey and Rentschler (1991) and Strasburger, Rentschler and Harvey (1994) measure the contrast threshold for the identification of numeric characters as a function of stimulus size and eccentricity. They notice that only for high contrast stimuli the minimum target size as a function of retinal eccentricity follows the predictions of cortical magnification theory.

*Institut für Medizinische Psychologie, Universität München, Goethestrasse 31, D-80336 München, Germany.
†To whom all correspondence should be addressed [*Email* martin@groucho.imp.med.uni-muenchen.de].

For low contrast stimuli the minimum contrast thresholds, reached by enlarging target digits in peripheral retinal loci, are never as low as those reached by targets seen in foveal vision. Consequently, there is no scaling function that could provide the perceptual equivalence between direct and indirect viewing conditions. This result confirms the original observation made by Aubert and Foerster (1857), that peripherally presented objects look *qualitatively* different even if their size is increased.

The recognition of numerical characters studied by Strasburger *et al.* involves patterns with an intrinsically two-dimensional signal variation (see Zetzsche & Barth, 1990). Thus it is impossible to represent the numerals along a single feature dimension such as orientation, length or curvature. Rather the classification process has to rely on a combination of several feature dimensions corresponding to categorical concepts or prototypes of classes (Watanabe, 1985). For habitual stimuli such as numerals prototype concepts can be assumed to exist prior to the experiment and to be stationary over time ("we know what the numeral '1' looks like"). From a more general perspective, however, pattern recognition originates in the evolution, or learning, of concepts which act as determinants of the classification behaviour later on. Classification and learning therefore must be regarded as interrelated phenomena. The aim of the present study is to compare classification processes for foveal and extrafoveal vision at this fundamental level of concept formation.

According to our experimental paradigm human observers are trained at different retinal locations to classify stimulus patterns (Caelli, Rentschler & Scheidler, 1987). More specifically, subjects are first repeatedly shown samples of each of several classes and instructed as to which class the samples belong. Thereafter they are required to classify samples to one of the classes. This implies that subjects do not know in advance what defines a particular class, or, in other words, that subjects do not have an a priori class concept. Rather this concept emerges during a process of supervised learning and classification, which is derived from a standard procedure used in digital signal processing to train pattern classifiers (Duda & Hart, 1973; Ahmed & Rao, 1975). The stimuli are compound Gabor signals defined in a two-dimensional feature space. Each of the pattern classes consists of a set of samples, or feature vectors, forming a cluster in feature space.

By employing this standard paradigm of pattern recognition, we recently were able to demonstrate how the concept of parametric Bayes classifiers known in technical pattern classification can be extended to allow the characterization of human behaviour in supervised learning and classification (Rentschler, Jüttner & Caelli, 1994). This model assumes that human classification behaviour is based on *internal* feature states which can be linked to *physical* feature vectors. Physical and internal feature states are coupled by additive stochastic error signals that can be estimated on the basis of the experimental classification data. Within the current parametric version of our Bayesian approach, classes are represented by distributions of feature vectors, and perceptual concepts of classes by the corresponding distributions of internal feature states. There are three major advantages of this technique: (1) it allows us to reconstruct the internal representations of classes observers are employing when doing the classification task; (2) it permits inferences about structure and dimensionality of the underlying internal feature space; (3) it provides a means to trace the temporal evolution of these representations during learning.

Although the scope of the following comparative evaluation is to study these aspects separately for foveal and extrafoveal vision, it should be remembered that visual perception probably relies on the cooperation of both modes of seeing as well. For instance, there is a long history of speculations about possible forms of foveal–extrafoveal interaction in maintaining visual stability across saccades (for a recent review of the various ideas see Bridgeman, van der Heijden & Velichkovsky, 1994). Our investigation therefore should also enlighten some of the constraints of such cooperative phenomena.

METHOD

Stimuli

Compound Gabor signals were generated in a 128×128 8-bit pixel format with a linear grey-level-to-luminance function. Intensity profiles were defined by

$$G(x,y) = L_0 + \exp\left\{-\frac{1}{\alpha^2}(x^2+y^2)\right\}(a\cos(2\pi f_0 x) + b\cos(2\pi 3 f_0 x + \phi)),$$

where L_0 determines the mean luminance, α the space constant of the Gaussian aperture, a the amplitude of the fundamental, b that of the third harmonic, and ϕ the phase angle of the latter. Thus, the two-dimensional images consisted of a fundamental cosine waveform and its third harmonic modulated by an isotropic Gaussian aperture which decayed to 1/e in 32 pixels.

Signal variation was restricted to b and ϕ. This allowed the use of a two-dimensional feature space with the Cartesian coordinates

$$\xi = b\cos\phi$$

("evenness") and

$$\eta = b\sin\phi$$

("oddness"). Three signal sets (I–III), each consisting of 15 samples, were used in the experiments (Fig. 1). The 15 signals of each set were grouped into three classes forming three clusters in feature space, each containing five samples. The three sets of signals differed in the grouping of their samples and, therefore, in the degrees of within and between variance in feature space.

The Gabor signals were displayed on a raster monitor (Lucius and Baer GBM 2310, P4 phosphor) linked to a digital image processing system (Matrox PIP 1024

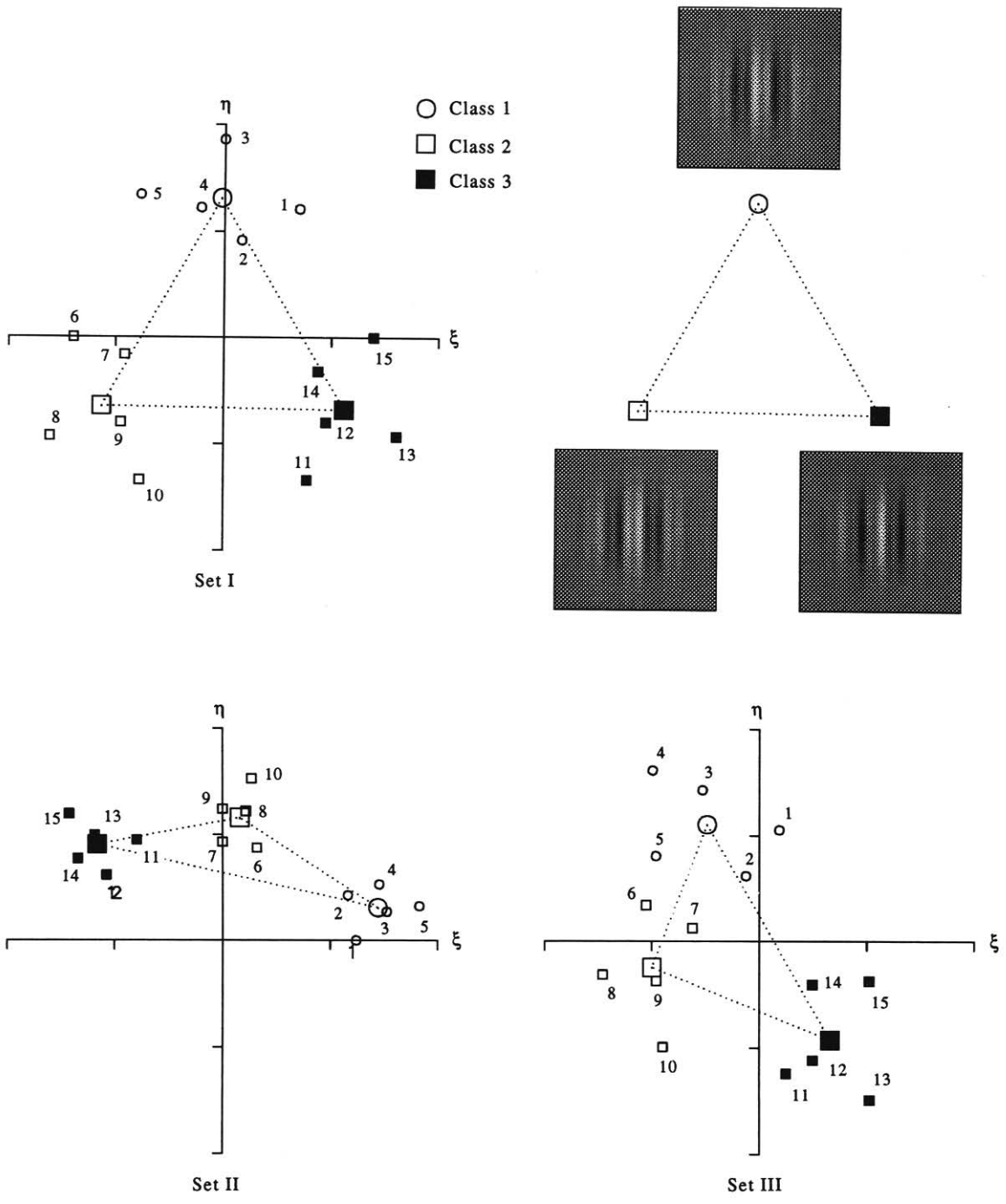

FIGURE 1. Feature space representations of the signal sets used in the experiments. Each of the three learning sets (I–III) consists of 15 compound Gabor patches. The fundamental frequency component is kept fixed in amplitude and (cosine) phase, whereas the third harmonic component varies in amplitude and phase. This allows the use of a metric feature space with the cartesian real (even), $\xi = b \cos \phi$, and imaginary (odd), $\eta = b \sin \phi$, coordinates (scale: 1 unit corresponds to 20 cd/m^2 amplitude relatively to mean luminance). The 15 signals of each learning set (small symbols) form three clusters in the feature space thus defining three classes. Large symbols refer to the class means, or physical prototypes. For set I the prototype signals of the three classes are depicted in the upper-left.

mounted on a PC-AT 286). Space average luminance (d.c.) was kept constant at 70 cd/m^2.

The stimulus patterns subtended 1.7 deg at a viewing distance of 101 cm when seen foveally. The fundamental spatial frequency was 2.4 c/deg. In 3 deg off-axis viewing conditions, the stimulus size was re-scaled to 2.7 deg according to cortical magnification (Rovamo & Virsu, 1979), i.e., by reducing the viewing distance by a factor of 1.6 (to 63 cm). Eccentricity was measured as the distance between the fixation point and the centre of the stimulus pattern, a definition corresponding to a fixation 1.65 deg away from the left or right edge of the target.

Procedure of supervised learning and classification

The learning procedure consisted of a variable number of learning units. One *learning unit* contained three

subsequent presentations, in random order, of the learning set with 200 msec exposure duration for each pattern. Following each presentation, a number specifying the class to which the pattern belonged was displayed for 1 sec. The interval between the learning signal and the number was 500 msec.

Each learning unit ended with a test of how well the subject was able to classify the 15 patterns. Only one exposure per sample was used here. The learning procedure was repeated for each observer until he or she had achieved an error-free classification (100% correct), or refused to continue due to an apparent inability to succeed with the task.

The procedure of supervised learning and testing was applied to each observer in three conditions: "central" (learning and testing in direct view), "left" (learning and testing when fixating a fixation target presented on the horizontal meridian 3 deg to the left of the centre of the stimulus patterns) and "right" (same but on the right side). Hence, in *all* conditions learning and testing occurred at the *same* retinal location. Five of the subjects started with the "central" classification task before switching to the eccentric conditions "left" and "right". The other seven performed the experiments in the reverse order, i.e., starting with the eccentric conditions. Viewing was always binocular.

Subjects

Twelve paid observers participated in the experiments, nine of which had never participated in any psychophysical experiment before. The remaining three subjects had extensive experience with other tasks involving peripheral vision. The ages of the 12 observers ranged between 20 and 30 yr, seven of them were female, five were male; all had normal or corrected-to-normal vision.

Bayesian model for supervised learning and classification

Our classification model (see Appendix I and Rentschler *et al.*, 1993, 1994 for a more detailed description) is based on the parametric Bayesian approach of technical pattern recognition (e.g. Duda & Hart, 1973; Ahmed & Rao, 1975). Specifically, it rests on three assumptions concerning how (1) a given physical stimulus is internally (perceptually) represented in the observer (signal representation assumption), (2) the representations of different stimuli belonging to the same signal class are distributed (class representation assumption) and (3) this distribution relates to the probability that a given signal will be attributed by the observer into this signal class (classification assumption).

The signal representation assumption. A given physical stimulus in the external world, uniquely specified by a feature vector in an appropriate physical feature space, gives rise to a specific *perceptual realization*. In other words, perception is regarded as some sort of measurement process which acts on the physical signal vector and results into another signal vector, corresponding to the *internal representation* of the external stimulus. Both representations necessarily differ since the perceptual process of internal feature measurement introduces additional degrees of freedom of both bias and variance. Following the standard approach of estimation theory (e.g. Gelb, 1974) we model the relationship between physical and internal signal representation in terms of an additive error vector to account for this additional variability. This means the internal signal representation corresponds to the vector sum of physical signal vector and perceptual error vector—an idea that also underlies multidimensional extensions of signal detection theory (Ashby, 1988, 1992).

The class representation assumption. Signal vectors of stimuli belonging to the same stimulus class are assumed to be parametrically distributed. This means that the signal vectors in their feature space representation form clusters, the spread of which can be described by parametric distribution functions. There is indeed preliminary evidence that subjects enter categorization tasks with the expectation that the exemplars of each category are symmetrically and unimodally distributed around some prototype (Fried & Holyoak, 1984; Flannagan, Fried & Holyoak, 1986). For the sake of mathematical simplicity, we assume all distributions to be of the Gaussian type. Now, given that both the physical signal vectors and the perceptual error vectors are normally distributed, the sum of both vectors—corresponding to the internal signal representation—will be normally distributed as well, due to the additivity of Gaussians, and its mean and covariance will be given by the vector sum of the means and covariances of the individual summands.

The classification assumption. Given the stochastic nature of the internal representation of individual signals and signal classes, the probability of a given signal to be assigned to a particular class is then provided by the Bayes theorem. In the present case (i.e., equal a priori probabilities assumed) it reduces to the value of the probability class density function of that particular class for the input signal, divided (normalized) by the sum of the density functions of all classes. This expression is formally similar to the similarity-choice models of Shepard (1957) and Luce (1963). However, it explicitly states that the characteristics of human classification behaviour do not directly depend on physical input, but on its corresponding internal representation.

The above assumptions form the basis of a *probabilistic virtual prototype (PVP)* classification model with four parameters per class, namely the mean values and covariances of the distributions of the physical input signals and of the perceptual error vectors, respectively. The former two are already completely specified by the distributions of the signals in physical feature space. The mean and the covariance of the error vector distribution, however, are free parameters of the model: they capture the additional degrees of *bias*, or distortion, and that of *covariance*, or fuzziness, of the internal representation for a particular class. The class means of the input signal distributions can be interpreted as class prototypes in

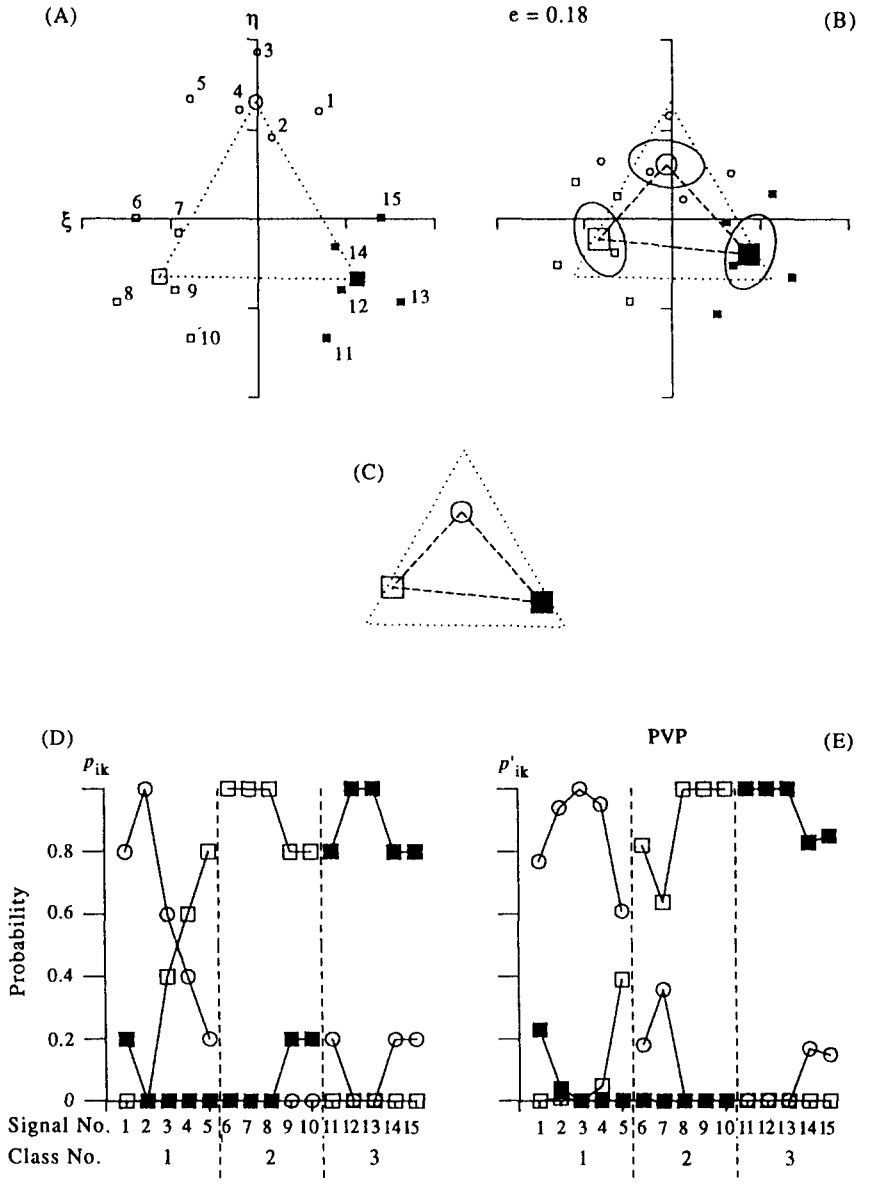

FIGURE 2. Classification of compound Gabor signals of learning set I. (A) Signal configuration, replotted from Fig. 1. (B) Virtual prototype solution (large symbols) obtained by minimizing the mean squared error e between observed classification data (D) and predicted classification data (E). Note that, according to the assumptions of the bias model, only the class means (or virtual prototypes) are displaced relative to the configuration of their physical counterparts (dashed triangle) whereas the clusters themselves and therefore the cluster covariances (indicated by ellipses) remain unaltered. (C) Graphic simplification of (B), to be used in Figs 3–6. (D) Data diagram showing, for each signal, the observed frequency of a classification into class 1 (○), class 2 (□), and class 3 (■). (E) Classification probabilities were predicted by the Bayesian classifier model.

physical feature space. Their counterparts in the internal representation may then be regarded as prototypes in a *virtual* sense: although physically not existing, the *virtual prototypes* allow us to account for the observed classification behaviour of human observers by a Bayesian classifier operating on the corresponding biased density functions.

Within the PVP approach, the mean and the covariance of the perceptual error distributions are the essential parameters capturing the internal representations of the signal classes underlying the classification judgement. In principle, both types of parameters can be estimated by minimizing the mean-squared error between the predicted classification probabilities and the experimentally observed classification probabilities. For the analysis of the experiments described in the following section, we make the further assumption that the covariances are identical for all classes. They can then be set arbitrarily to zero. In other words, we assume that the class conditional probability densities in physical and internal feature space have the same covariances and differ only in their mean values, or in their class specific bias (bias model). This restriction is due to the technical limitations of our experimental set-up, which allows the use of random sequences of 15 signals with 128×128 pixels at most. This would seem to be a set of samples that is too small for reasonably robust simultaneous estimates of *both* mean and covariance of the perceptual error distributions

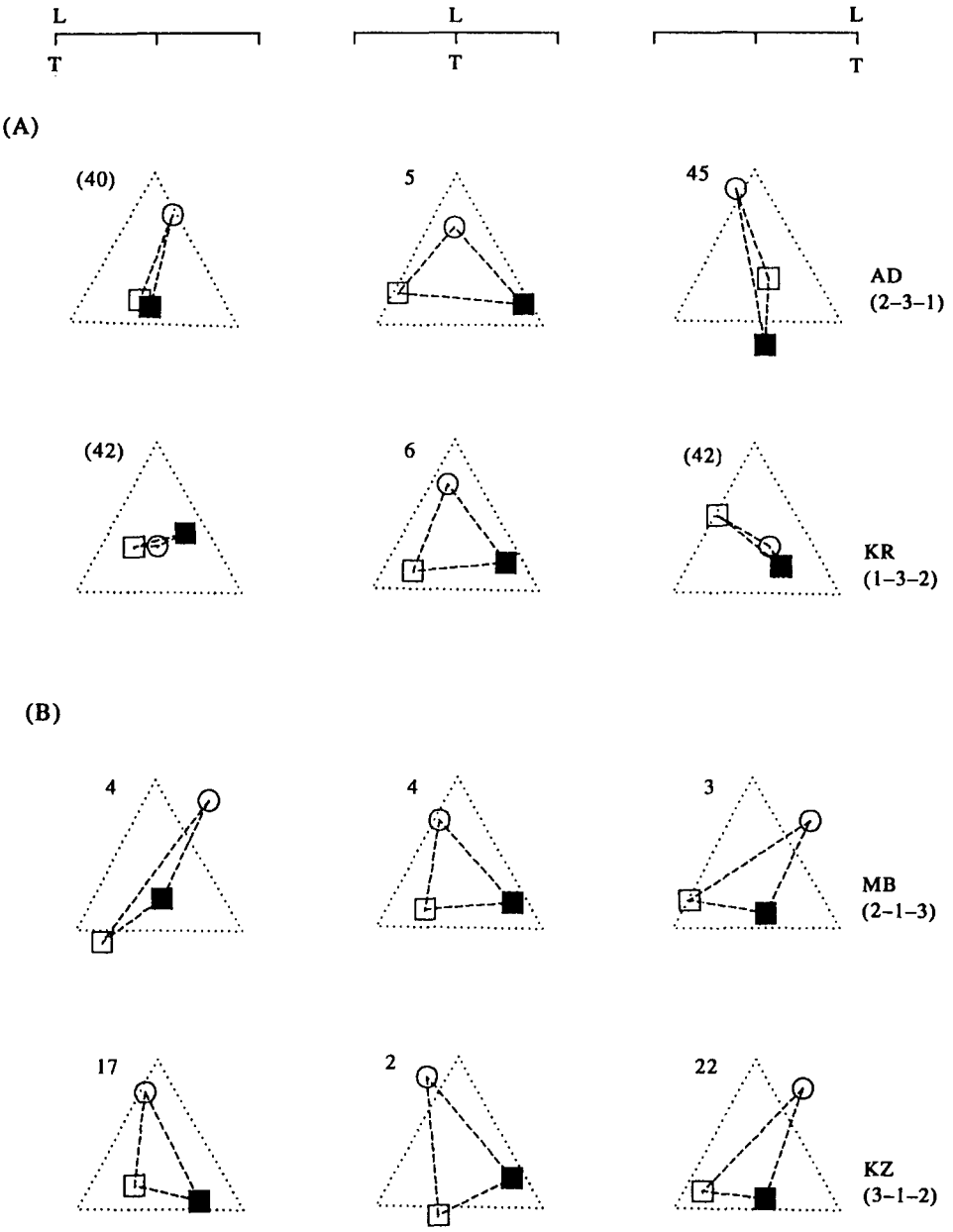

FIGURE 3. Supervised learning and classification of signal set I in direct and indirect view for two naive (A) and two experienced (B) observers. Each row shows the virtual prototypes estimated from the classification data of one observer for condition "left", "central" and "right". Note that compared to the configuration of the physical prototypes (dotted triangle), the configuration of the virtual prototypes appears degenerated to a nearly collinear formation in eccentric viewing for naive subjects. For each condition the number of learning units is listed. Numbers in parentheses refer to cases where the subject stopped the experiment without having reached a perfect classification. The order in which the different conditions (left–central–right) were tested is also given.

[the bias/variance dilemma of statistical inference, cf. Geman, Bienenstock and Doursat (1992)].

The procedure outlined above may not only be applied to classification data cumulated across the entire procedure of visual learning. Rather it is possible (cf. Appendix I to trace the evolution of internal representations (visualized by the configurations of the virtual prototype configurations) by using temporally constrained estimates of the classification probabilities. This results into a series of virtual prototype configurations, the so-called *learning tomograms*, that allow us to account for the dynamics of the learning process.

RESULTS

Classification performance of naive observers: foveal and extrafoveal vision

Before presenting the results of the different experimental conditions we first illustrate the application of the PVP approach by discussing the results of one subject in greater detail. Figure 2(A) shows the samples of learning set I (replotted from Fig. 1) that were to be classified by the observer (AD, central viewing). The signals form three clusters in feature space thus defining three classes. The positions of the three *class means*, or physical

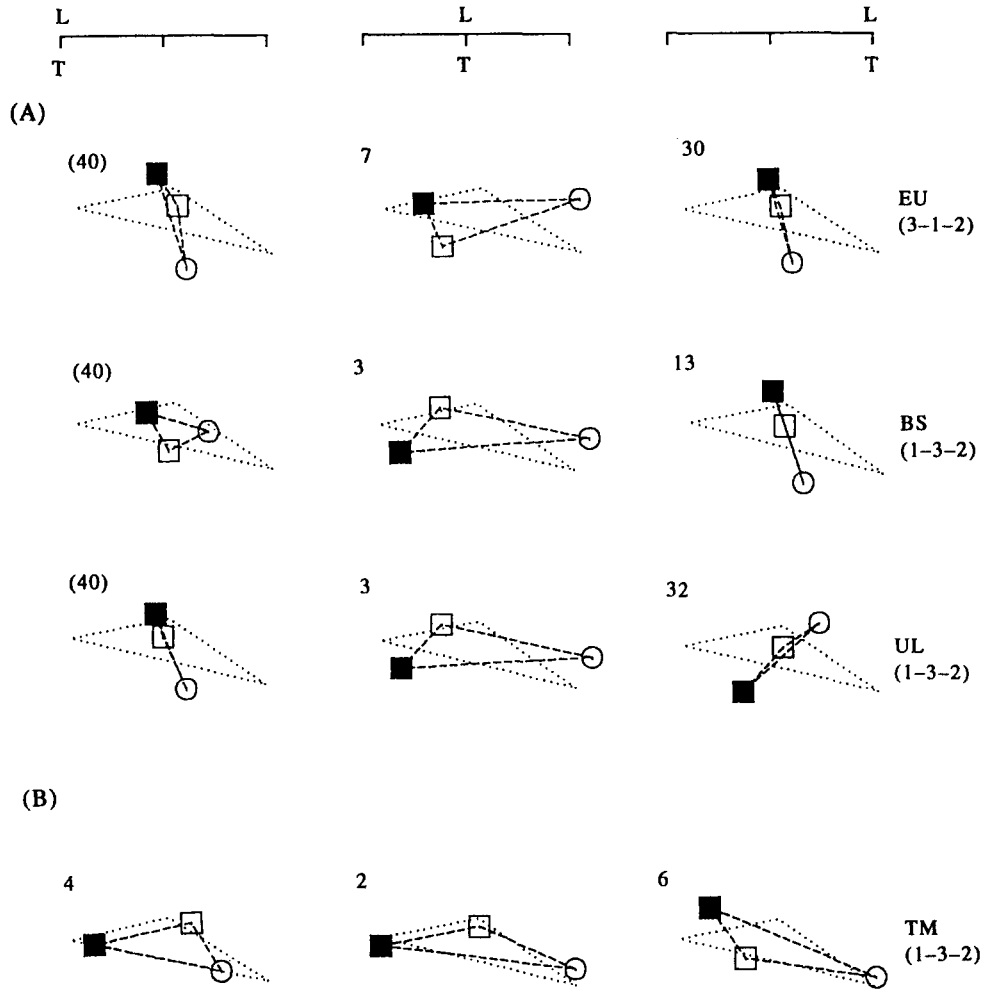

FIGURE 4. Supervised learning and classification of signal set II in direct and indirect view for two naive (A) and one experienced (B) observer. See Fig. 3 for further explanation.

prototypes, are indicated by the large symbols at the vertices of the dashed triangle. Below, in Fig. 2(D), the experimental classification data are represented, i.e. the probability of a given signal (1–15) being assigned to class 1 (○), class 2 (□) and class 3 (■) by the observer.

Figure 2(B) shows the three *virtual prototypes* obtained by employing the parametric Bayesian classifier on the corrected learning signals and minimizing the mean-squared error between observed [Fig. 2(D)] and predicted classification probabilities. The resulting classification probabilities of the model are shown in Fig. 2(E). This optimum solution has been obtained under the assumption of the *bias* model. This means that the signal clusters are displaced in feature space with translations (the added error vectors) only, thus leaving the cluster covariance matrices [represented by the ellipses in Fig. 2(B)] fixed. Consequently, the signal clusters do not undergo rotations during the optimization procedure. Note that, for a better visualization, virtual prototypes are re-mapped into physical feature space although they describe *internal* representations. For the sake of clarity, virtual prototype configurations will be reproduced in a simplified version [Fig. 2(C) instead of (B)].

Figure 2 is based on the data of an experiment where the subject saw the patterns in direct view during supervised learning and testing. The relative positions of the virtual prototypes with respect to each other are similar to those of the physical class means marked by the vertices of the dashed triangle in Fig. 2(B, C). The fact, that both configurations are not exactly congruent to each other indicates the presence of some bias in the internal representation. The existence of bias, however, presents no obstacle for correct classification.

Virtual prototype configurations for foveal and extrafoveal view are shown in Fig. 3 for subject AD and three other observers (KR, MB and KZ). Figure 3(A) shows the reconstructed virtual prototype configurations for two naive subjects and the same set of signals as used for Fig. 2. On the left of the figure are data, which were obtained at a condition where learning and testing was done while fixating a fixation target 3 deg to the left of the stimulus patterns. On the extreme right of Fig. 3(A) the corresponding results for the complementary task, e.g. fixation at 3 deg to the right, are depicted. Compared to the case of central viewing [results plotted in the middle part of Fig. 3(A)], the virtual prototypes appear severely distorted in the sense that their formation is degenerated

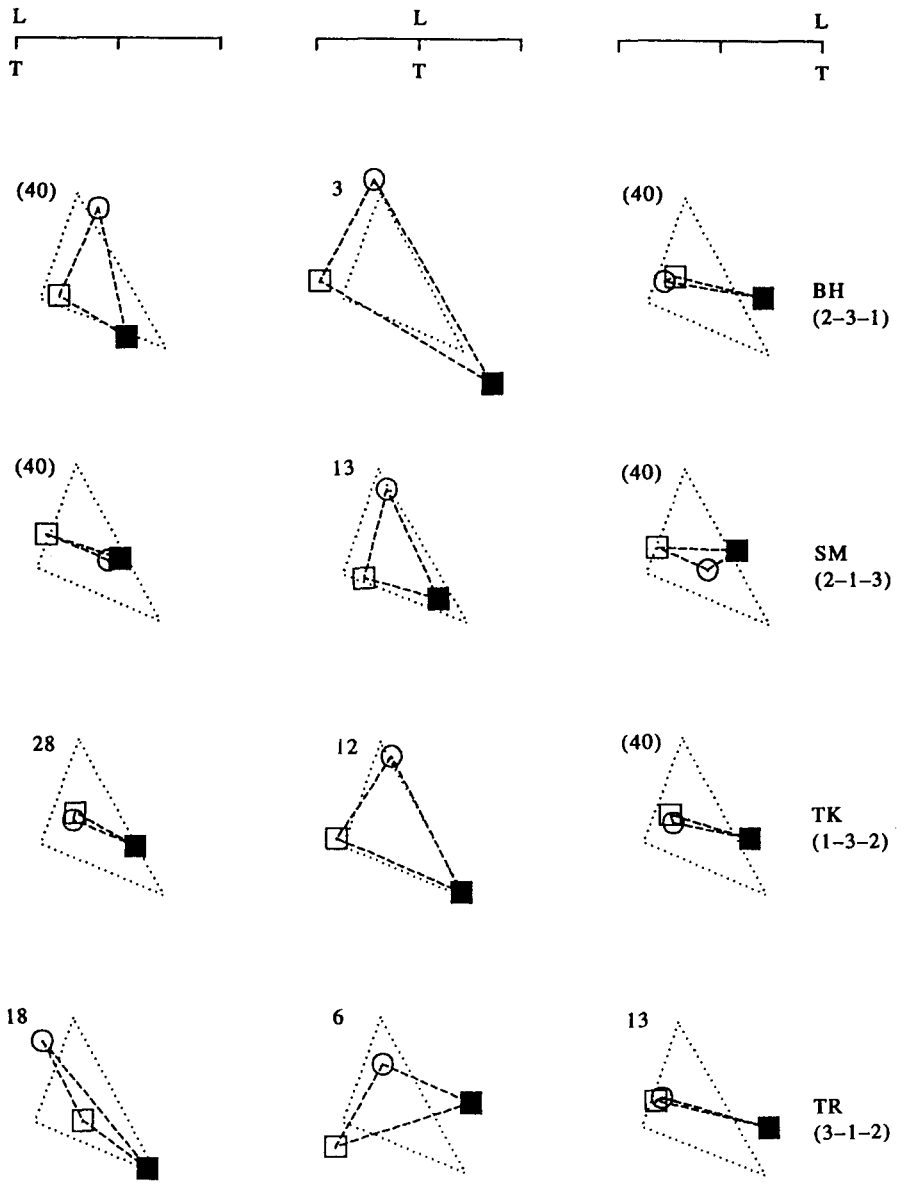

FIGURE 5. Supervised learning and classification of signal set III in direct and indirect view for four naive observers. See Fig. 3 for further explanation.

to a quasi one-dimensional configuration. This characteristic is found *although* all signals were *size scaled* according to the cortical magnification concept. Nevertheless, size scaling does not yield a renormalization of the internal representation. This is even more conspicuous if one considers the number of learning steps necessary to achieve perfect classification. As shown in Fig. 3(A) about 40 learning units are required for naive observers; moreover, there are some cases (numbers included in parentheses) where subjects refused to continue learning without having reached the criterion (100% correct).

The differences between foveal and extrafoveal vision demonstrated by reconstructing the underlying internal representations via the PVP approach are not dependent on the very choice of the set of signals used in the learning experiment. Figures 4(A) and 5 show the results for the two other learning sets involving different signal configurations (dashed triangles) in physical feature space. These data were obtained with two other groups of naive observers (EU, BS, UL, TM, BH, SM, TK and TR). In the plots the locations of the physical class means (cf. the distributions depicted in Fig. 1) are indicated by the vertices of the dashed triangles. From a comparison of physical and virtual prototypes, it is again obvious that the high classification performance for *central* viewing is related to internal representations, which deviate relatively little from the objective ones, whereas *extrafoveal* learning and classification are based on strongly distorted signal representations.

Classification performance of observers without specific experience

The deficits of classification performance in eccentric viewing cannot, as has been already pointed out, be compensated for by size scaling. Yet, the data provide

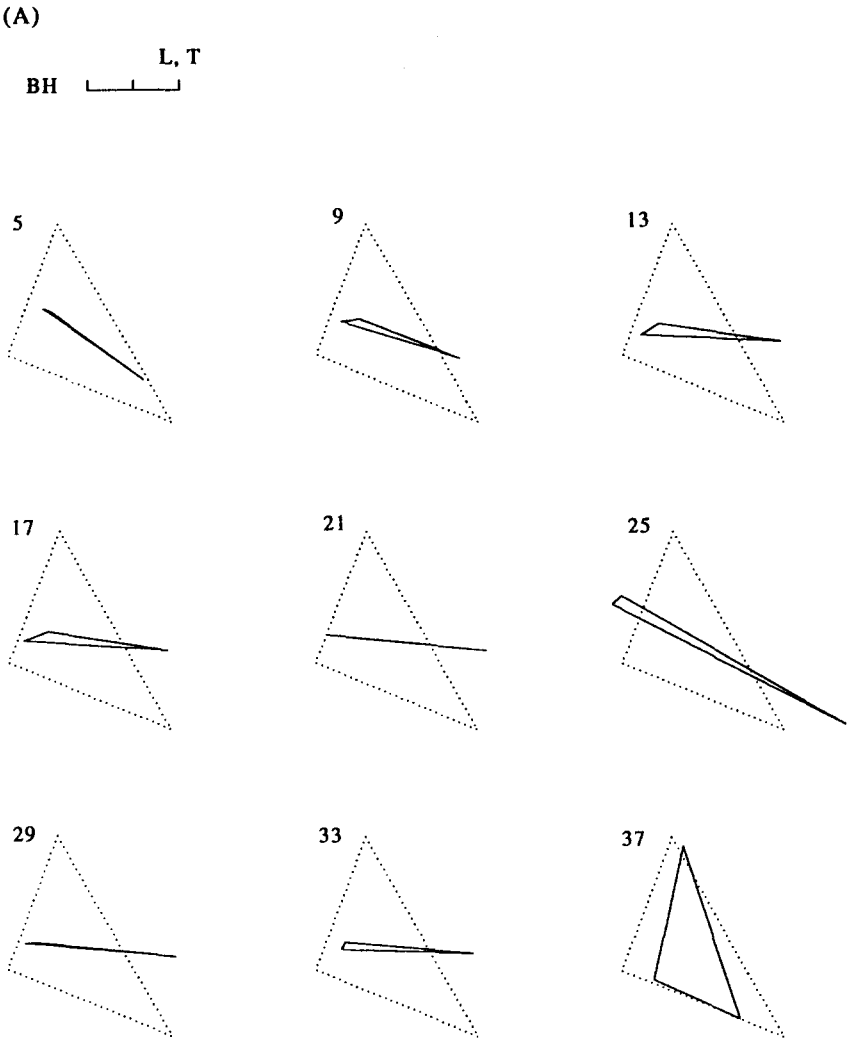

FIGURE 6. (A).

evidence as well that these deficits are not static but can be reduced by training. For instance, both subjects BS and UL [Fig. 4(A)], who started the experiment with condition "left" without reaching a perfect classification within 40 learning units, did better in the second task when switching to condition "right". Note, that these improvements in performance are obscured in Fig. 4(A) because the present plots are based on cumulative classification matrices resulting from temporal averaging (the dynamic aspects of learning will be discussed in the following section on "learning tomograms").

The improvements of classification performance via supervised learning have to be distinguished from the effects of *unspecific experience* with extrafoveal vision. Figures 3(B) and 4(B) plot learning data from three observers (MB, KZ and TM), who had never participated in classification experiments before. These subjects, however, had extensive experience with character recognition (MB and KZ; see Strasburger *et al.*, 1991, 1996) and double-pulse resolution (TM; see Treutwein & Rentschler, 1992) in the peripheral visual field. Compared with the corresponding plots of naive subjects [Figs 3(A) and 4(A)], the differences between the virtual prototype configurations of central and eccentric learning conditions are far less pronounced here. All of the unspecifically experienced observers reached the classification goal. While for foveal learning the number of learning units appears to be only marginally reduced with respect to the naive subjects, the improvement for extrafoveal learning is tremendous: actually two of the three subjects (who were in fact the most experienced ones) reached the criterion nearly as fast as naive subjects did in the central condition, i.e., within six learning units.

Our finding, that the properties of internal representations are obviously far from being constant over time, raises the question of their dynamic behaviour. This issue will be addressed in the following section.

Learning tomograms: the evolution of internal representations

The virtual prototypes considered so far were derived from the overall cumulative classification matrices reflecting the temporal average of classification performance. To capture the dynamics of the learning process it is useful to base the prototype computation on temporal

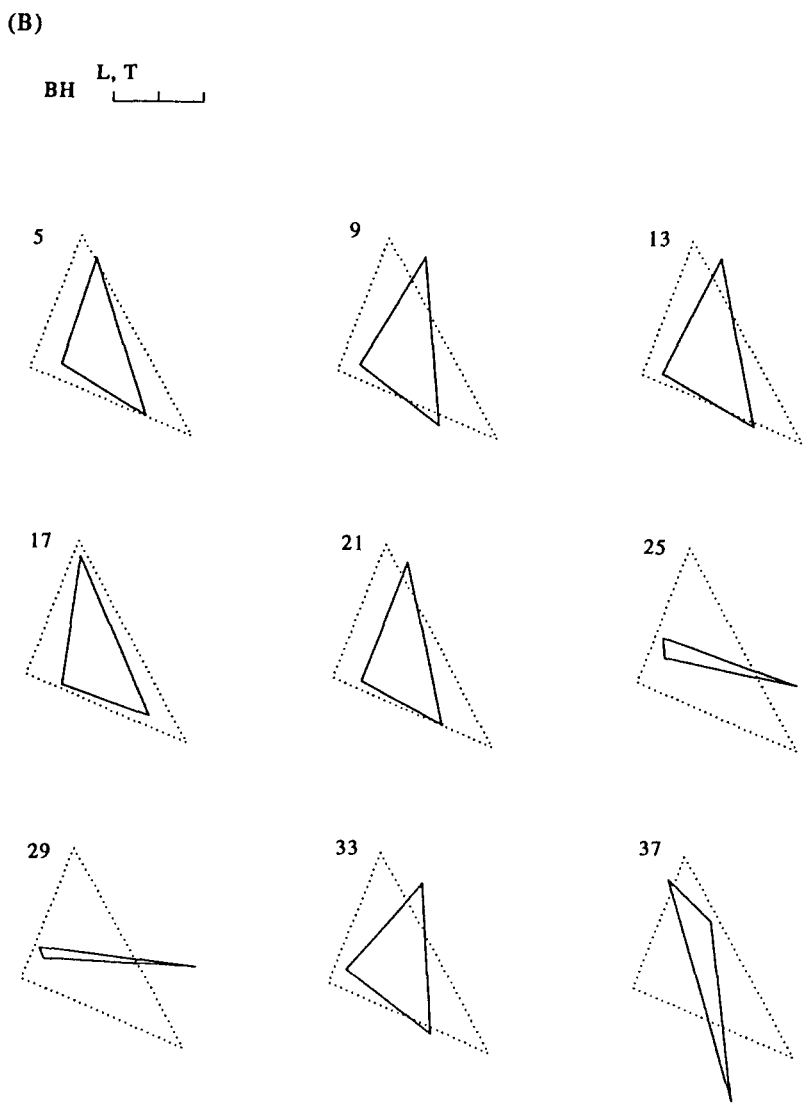

FIGURE 6. Learning tomograms of a naive observer for extrafoveal learning with (A) right eccentric and (B) left eccentric fixation during learning and testing. The tomograms were computed on the basis of classification matrices which were averaged over truncated temporal Gaussian windows. The width of the latter was 4 learning units. This means that, for instance, the estimate for the prototype configuration at trial 5 in (A) is based on the responses of trials 3–7. The configurations for trials 9, 13, 17 etc. were calculated in an analogue way.

windows of the classification matrix. This idea leads to a representation, where the learning process is traced by a series of temporally constrained estimates of prototypes, the learning tomograms. Figure 6 illustrates the procedure for the data of subject BH, who started the experiment with condition "right" [Fig. 6(A)] and then switched to condition "left" [Fig. 6(B)]. The learning tomograms are calculated from classification matrices, which were averaged over truncated Gaussian windows in the temporal domain. Here, the width h of the window function in A(9) was set to 4 learning units. This means that, for instance, the estimate for the prototype configuration at trial 5 is based on the responses from trials 3–7, while that at trial 9 is based on the responses of trials 7–11, and so on.

The series of tomograms thus demonstrates the temporal evolution of internal representations during learning. From Fig. 6 it becomes evident that learning does not appear as a smooth process of gradual development of adequate internal representations. Rather there are periods of quasi stationary prototype configurations [e.g. from trial 5 to 33 in Fig. 6(A), and from trial 5 to 21 in Fig. 6(B)], followed by abrupt transitions between certain reoccurring configurations [e.g. the tomogram sequence 21–25–29–33 in Fig. 6(B)]. This suggests that visual learning may be characterized more appropriately by the reoccurrence of certain *stereotypes*. In Fig. 6(A) one stereotype may be assigned to the prototype configurations estimated for instance at trials 5, 9 and 13, and another one to that at trial 37. In Fig. 6(B) the stereotypes would correspond to the configurations estimated at trials 5, 25 and 37. Interestingly, the first two stereotypes of Fig. 6(B) are quite similar to those appearing in Fig. 6(A) [cf., e.g. the stereotype of trial 5

in Fig. 6(B) with that of trial 37 in Fig. 6(A), or that of trial 25 in Fig. 6(B) with the one estimated at trial 13 in Fig. 6(A)]. It is therefore intriguing to assume that the alternation between specific types of configurations corresponds to the existence of different *working hypotheses* or class concepts which the subject puts forward to succeed in the classification task. Each of these competing hypotheses manifests itself in a different pattern of correct and incorrect classifications and, therefore, in a specific configuration of virtual prototypes.

Furthermore, the fact that the last stereotype under condition "right" [Fig. 6(A), trial 37] reappears as the first one under the subsequent condition "left" [Fig. 6(B), trial 5] suggests that the learning progress achieved at one site in the visual field is in part transferred to the contralateral side—a fact that is obscured in the virtual prototype configurations derived from the cumulative classification data (Figs 3–5). Additional evidence for such a contralateral transfer of performance is provided by a statistical analysis comparing the cumulated classification performance of learning units 1–5 of the first extraretinal learning condition with that of the second extraretinal learning condition (note that all subjects performed the extraretinal conditions in direct succession). This analysis was performed for all nine naive subjects. All of them show an improvement of performance in the second condition, which in six cases reaches significance in a paired t-test with $\alpha \leqslant 0.05$.

Finally, it should be noted that we have restricted our discussion of the learning tomograms to the case of peripheral learning. The reason for this is the fact that for central viewing conditions observers in our experiments usually achieved perfect classification within 5–7 learning units. The series of learning tomogram representations then reduces to a one-step representation which is practically identical to the cumulative one shown in Figs 3–5.

DISCUSSION

We have shown that human pattern recognition of complex stimuli differs considerably between foveal and extrafoveal viewing conditions. The difference became especially evident for naive subjects in that they needed a much longer learning phase for extrafoveal learning than for foveal. Some of these subjects even never succeeded in achieving perfect (100% correct) classification. If they did, the corresponding class concepts, as visualized by virtual prototypes, revealed strong distortions relatively to those of foveal learning, which were more directly related to the corresponding physical prototypes. For unspecifically experienced observers this functional dichotomy was less conspicuous, though learning also proceeded more slowly in the extrafoveal condition.

The differences between foveal and extrafoveal classification behaviour emerged despite of the fact that the image signals were scaled according to cortical magnification theory. This result is consistent with the findings of Strasburger *et al.* (1994) who measured the recognition of characters in direct and eccentric view. They observed that only the recognition of high contrast characters is accommodated by the cortical magnification concept, whereas the recognition of low contrast characters is not. Strasburger *et al.* attributed this finding to the fact that the recognition of complex (intrinsic dimension >1D) stimuli requires the combination of several features, each of which introduces its own threshold in the sense of a discriminant function in feature space (Watanabe, 1985). The recognition process thus requires a feature ponderation along *all* relevant stimulus dimensions. Consequently, size scaling must remain insufficient, unless the other relevant dimensions are sufficiently represented.

The present study differs from that of Strasburger *et al.*, since the latter employed a postcategorical recognition task, where subjects were a priori familiar with the categories at issue (i.e., the numerical characters 0–9). We studied the process of visual recognition on a more general, precategorical level, in that categories had to be learned during the experiments. Under such conditions, two steps have to be distinguished: (1) relevant feature dimensions for the classification task have to be established and (2) feature weights (i.e., coordinates in feature space) need be determined.

As to the feature representation used in this study, the following *caveat* is necessary. The choice of a two-dimensional feature space of even- and oddness dimensions does not imply that such features are necessary or even sufficient for human pattern recognition in general. Yet it is plausible due to the notion that pairs of even and odd symmetric filters centred at the same location serve as basis functions of spatial vision (e.g. Robson, 1975; Pollen & Ronner, 1981). It is also useful since it yields a parsimonious signal description with pattern classes forming clusters in feature space. However, it is in principle impossible to prove uniqueness of a given feature representation for solving a pattern recognition problem, so that its plausibility has to be judged with respect to additional criteria (see Watanabe, 1985). Thus it is possible that the outputs of evenness and oddness filters do not directly underlie visual pattern recognition but are first re-combined in a nonlinear fashion. Related theoretical schemes are local energy (Burr, Morrone & Spinelli, 1989) and localized phase (Zeevi & Porat, 1989), or local amplitude and phase (Wegmann & Zetzsche, 1990).

Common to the aforementioned concepts of visual pattern processing is the notion of feature localization, i.e., of linking features to specific image parts. Rentschler and Treutwein (1985) have shown that there exists a scale-invariant weakness of extrafoveal vision in the processing of positional information, and it is also known that related deficits of local feature extraction can be amended by directing focal attention (Beck & Ambler, 1973; Bergen & Julesz, 1983). Such strategies, however, require extensive training. This would explain the advantage of our semi-experienced subjects, who had extensive experience with character recognition (MB, KZ; see Strasburger *et al.*, 1991, 1996) and double-pulse

resolution (TM; see Treutwein & Rentschler, 1992) in the peripheral visual field, although they were completely naive with respect to the present paradigm.

The characterization of internal representations by their perceptual dimension

The distinct differences between foveal and extrafoveal pattern recognition manifest themselves in the structure of virtual prototype configurations. The internal representations of the physical class means display, in the case of peripheral pattern presentation, a nearly one-dimensional arrangement. Such distortions suggest a restriction of the degrees of freedom characterizing internal representations. The mathematical concept of *dimension* provides an appropriate means to make the notion of degree of freedom more explicit, a notion which also applies to its most general formulation, the concept of fractal dimension (Farmer, 1982). So far, fractal analysis has been applied to human vision mainly in the context of roughness perception (Pentland, 1984), shape perception of patterns with fractal contours (Rogowitz & Voss, 1990), and curvature representations on multiple scales (Barth, Zetzsche, Ferraro & Rentschler, 1993). It should be noted, however, that the term "fractal dimension" has no unique definition, although it is often identified with the Hausdorff dimension or capacity dimension. Rather, it is a generic name for dimensions which can take fractional values (Takayasu, 1989).

On this basis, we show in Appendix II that it is possible to derive a *perceptual dimension* D_ψ which may be used to characterize internal representations within the PVP approach. This dimension is *not* a fractal dimension in a strict mathematical sense (however, it should be annotated that this holds true for all applications of fractal geometry in the physical world). The essential point is that we adopt, for a given scaling range confined by the upper boundary of the between-class distance and the lower boundary of the within-class variance, the capacity definition from fractal analysis to define a dimensional index for the configurations of the virtual prototypes. Although the notion of a *perceptual* dimension cannot, strictly speaking, apply for signals defined in a *physical* feature space, we analogously compute this index (D_ϕ) also for the physical prototype configurations. We thus provide a standard of reference in order to make the structural differences between physical and internal signal representations more explicit. Table 1 contrasts the perceptual dimensions for the virtual prototype configurations depicted in Figs 3–5 with the corresponding indices for the physical prototypes. For foveal learning the perceptual dimension of the internal representation matches that of the underlying physical class means, i.e., the internal representation provides a *complete* description of the physical feature space with respect to the classification task. For peripheral learning, however, the perceptual dimension for naive subjects is clearly reduced to values about unity corresponding to an *incomplete* representation. By contrast, the data for experienced observers do not reveal this sharp decline. Here the perceptual dimensions for both foveal and extrafoveal viewing conditions remain comparable.

TABLE 1. Perceptual dimension D_ψ of the virtual prototype configurations shown in Figs 3–5

Subject	D_ψ of virtual prototype configuration		
	Cond. left	Cond. central	Cond. right
	Learning set I (D_ϕ=1.7)		
AD	1.1	1.8	1.2
KR	1.0	1.8	1.1
MB	1.3	1.8	1.5
KZ	1.5	1.6	1.6
	Learning set II (D_ϕ=1.4)		
EU	1.0	1.4	1.0
BS	1.2	1.5	1.1
UL	0.9	1.5	0.9
TM	1.3	1.4	1.3
	Learning set III (D_ϕ=1.7)		
BH	1.5	1.6	1.0
SM	1.1	1.5	1.2
TK	1.0	1.6	1.1
TR	1.3	1.6	1.1

The values in parentheses refer to the corresponding dimension D_ϕ of the physical prototype configuration for the underlying learning set.

The fact that in the plots of Figs 3–5, the configurations of virtual and physical prototypes in central viewing are not congruent demonstrates that a perfect classification does not imply the internal representation to be bias-free. Again, the basic criterion guaranteeing the adequate separation of the class representatives is that of completeness. In other words, it is the *perceptual dimension* of the signal configurations that has to be preserved in the internal representation, not their structure *per se*.

The fall-off of visual performance with increasing retinal eccentricity is known to be distinctly task-dependent (cf. Wilson *et al.*, 1990). For instance, the rate of eccentricity variation for vernier acuity differs from that of grating acuity by nearly a factor of 3 (Levi *et al.*, 1985). The question therefore arises whether the deficits obtained under extrafoveal learning conditions could have been compensated for by rescaling the signals with a different scaling factor for pattern size. The analysis of our data in terms of a perceptual dimension strongly argues against this possibility. From the results listed in Table 1 it seems to be clear that pattern recognition ability in direct and indirect view differ in a more fundamental respect than can be accounted for by any linear concept. Rather these results indicate a *dimensional incommensurability* of the underlying representations. Furthermore, our observation that the deficits of extrafoveal vision are not static but can be amended in part by learning appears to be incompatible with the idea that they are related to a neuroanatomical constraint provided by the retino-cortical mapping.

Implications for foveal–extrafoveal cooperation

On the one hand we have shown that the internal representations of foveal and extrafoveal vision are intrinsically different with respect to their perceptual dimensions. Hence they are, strictly speaking, incommensurable. On the other hand, ecological perception requires the combination of information originating from different parts of the visual field. During normal explorative activity of the eye, parts of the visual field containing objects of interest become successively mapped on both extrafoveal and foveal parts of the retina. This visual scanning of the environment is mediated by saccadic eye movements during which sensitivity is greatly reduced because of saccadic suppression. Nevertheless we experience the world stable and continuous rather than being fragmented into individual glimpses.

A classical explanation of this phenomenon of visual stability is that the perceptual system constructs a representation based on spatiotopic coordinates. Different versions of this type of solution exist (see e.g. Bischof & Kramer, 1968; Breitmeyer, Kropfl & Julesz, 1982; Feldman, 1985) but their common ratio is the notion of some sort of iconic memory where spatiotopic recalibrated versions of the retinal images originating from successive fixations superimpose each other. Such an integration across saccades obviously would explain perceptual continuity. However, psychophysical evidence has disproved this hypothesis in numerous studies (e.g. Bridgeman & Mayer, 1983; Irwin, Yantis & Jonides, 1983; Jonides, Irwin & Yantis, 1983; Rayner & Pollatsek, 1983).

The failure of the integration solution appears plausible in the light of the present findings: it would imply a superposition of images which originate from foveal and extrafoveal parts of the retina and therefore are intrinsically incommensurable due to their different perceptual dimension. But this only constrains possible forms of cooperative foveal–extrafoveal mechanisms, not excludes them. The present experiments focused on learning conditions where learning and testing always occurred at the same (foveal or extrafoveal) retinal location. However, our analysis of the dynamics of learning revealed that certain stereotypes, characterizing internal concepts of classes, may be transferred across subsequent learning conditions, and therefore, across different retinal locations (cf. Fig. 6). Even if these findings are still preliminary and have to be extended by further investigations, they suggest that cooperative processes between foveal and extrafoveal vision may be restricted to a more abstract level. Indeed, there is some evidence for such a postcategorical interaction (O'Regan & Lévy-Schoen, 1983; Irwin, Brown & Sun, 1988; Pollatsek & Rayner, 1991; Jüttner & Röhler, 1993) though the level of abstraction has yet to be determined.

REFERENCES

Ahmed, N. & Rao, K. R. (1975). *Orthogonal transforms for digital signal processing*. Berlin: Springer.

Ashby, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics, 44*, 195–204.

Ashby, F. G. (1992). Multidimensional models of categorization. In Ashby, F. G. (Ed.), *Multidimensional models of perception and cognition* (Chap. 16). Hillsdale, N.J.: Lawrence Erlbaum.

Aubert, H. & Foerster, R. (1857). Beiträge zur Kenntnis des indirekten Sehens. I. Über den Raumsinn der Retina. *Albert von Graefes Archiv für Ophthalmologie, 3*, 1–37.

Barth, E., Zetzsche, C., Ferraro, M. & Rentschler, I. (1993). Fractal properties from 2D-curvature on multiple scales. In Vemuri, B. (Ed.), *Geometric methods in computer vision II, Proceedings SPIE, 2031*, 87–99.

Beck, J. & Ambler, B. (1973). The effects on concentrated and distributed attention on peripheral acuity. *Perception & Psychophysics, 14*, 225–230.

Bergen, J. R. & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature, 303*, 696–698.

Bischof, N. & Kramer, E. (1968). Untersuchungen und Überlegungen zur Richtungswahrnehmung bei willkürlichen sakkadischen Augenbewegungen. *Psychologische Forschung, 32*, 185–218.

Breitmeyer, B. G., Kropfl, W. & Julesz, B. (1982). The existence and role of retinotopic and spatiotopic forms of visual persistence. *Acta Psychologica, 52*, 175–196.

Bridgeman, B. & Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bulletin of the Psychonomic Society, 24*, 285–286.

Bridgeman, B., van der Heijden, A. H. C. & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences, 17*, 247–258.

Burr, D. C., Morrone, M. C. & Spinelli, D. (1989). Evidence for edge and bar detectors in human vision. *Vision Research, 29*, 419–431.

Caelli, T., Rentschler, I. & Scheidler, W. (1987). Visual pattern recognition in humans. I. Evidence for adaptive filtering. *Biological Cybernetics, 57*, 233–240.

Chandler, J. P. (1969). STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science, 14*, 81–82.

Cowey, A. & Rolls, E. T. (1974). Human cortical magnification factor and its relation to visual acuity. *Experimental Brain Research, 21*, 447–454.

Daniel, P. M. & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *Journal of Physiology, 159*, 203–221.

Duda, R. O. & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: John Wiley.

Farmer, J. D. (1982). Information dimension and the probabilistic structure of chaos. *Zeitschrift für Naturforschung, 37*, 1304–1325.

Feder, J. (1988). *Fractals*. New York: Plenum Press.

Feldman, J. A. (1985). Four frames suffice: A provisional model of vision and space. *Behavioral and Brain Sciences, 8*, 265–289.

Flannagan, M. J., Fried, L. S. & Holyoak, K. J. (1986). Distributed expectations and the induction of category structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 241–256.

Fried, L. S. & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 234–257.

Gauss, C.F. (1809). *Theoria motus*. English translation (1963). *Theory of the motion of the heavenly bodies about the sun in the conic sections*. New York: Dover.

Gelb, A. (1974). *Applied optimal estimation*. Cambridge, MA.: MIT Press.

Geman, S., Bienenstock, E. & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation, 4*, 1–58.

Harvey, L. O. Jr, Rentschler, I. & Weiss, C. (1985). Sensitivity to phase distortions in central and peripheral vision. *Perception & Psychophysics, 38*, 392–396.

Hering, E. (1899). Über die Grenzen der Sehschärfe. *Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wis-*

senschaften zu Leipzig, Mathematisch-Physische Classe; Naturwissenschaftlicher Teil (pp. 16–24).

Hilz, R. & Cavonius, C. R. (1974). Functional organization of the peripheral retina: Sensitivity to periodic stimuli. *Vision Research, 14*, 1333–1337.

Irwin, D. E., Brown, J. S. & Sun, J. S. (1988). Visual masking and visual integration across saccadic eye movements. *Journal of Experimental Psychology: General, 117*, 276–287.

Irwin, D. E., Yantis, S. & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics, 34*, 49–57.

Jonides, J., Irwin, D. E. & Yantis, S. (1983). Failure to integrate information from successive fixations. *Science, 222*, 188.

Jüttner, M. & Röhler, R. (1993). Lateral information transfer across saccadic eye movements. *Perception & Psychophysics, 53*, 210–220.

Kelly, D. H. (1984). Retinal inhomogeneity: I. Spatiotemporal contrast sensitivity. *Journal of the Optical Society of America A, 1*, 107–113.

Koenderink, J. J., Bouman, M. A., Bueno de Mesquita, A. E. & Slappendel, S. (1978). Perimetry of contrast detection thresholds of moving spatial sine wave patterns. III. The target extent as a sensitivity controlling parameter. *Journal of the Optical Society of America A, 68*, 854–860.

Levi, D. M., Klein, S. A. & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research, 25*, 963–977.

Luce, R. D. (1963). Detection and recognition. In Luce, R. D., Bush, R. R. & Galanter, E. (Eds), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.

Ludvigh, E. (1941). Extrafoveal visual acuity as measured with Snellen test letters. *American Journal of Ophthalmology, 24*, 303–310.

Mandelbrot, B. B. (1982). *The fractal geometry of nature*. New York: Freeman.

O'Regan, J. K. & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research, 23*, 765–768.

Pentland, A. P. (1984). Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 661–674.

Pollatsek, A. & Rayner, K. (1991). What is integrated across fixations? In Rayner, K. (Ed.), *Eye movements and visual cognition: Scene perception and reading*. New York: Springer.

Pollen, D. A. & Ronner, S. F. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science, 212*, 1409–1411.

Rayner, K. & Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception & Psychophysics, 34*, 39–48.

Rentschler, I., Jüttner, M. & Caelli, T. (1993). Ideal observers for supervised learning and classification. In Steyer, R., Wender, K. F. & Widamann, K. F. (Eds), *Psychometric methodology* (pp. 440–445). Stuttgart: Fischer.

Rentschler, I., Jüttner, M. & Caelli, T. (1994). Probabilistic analysis of human supervised learning and classification. *Vision Research, 34*, 669–687.

Rentschler, I. & Treutwein, B. (1985). Loss of spatial phase relationships in extrafoveal vision. *Nature, 313*, 308–310.

Robson, J. G. (1975). Receptive fields: Neural representation of the spatial and intensity attributes of the visual image. In Carterette, E. C. & Friedman, M. P. (Eds), *Handbook of perception* (Vol. 5, pp. 81–116). New York: Academic Press.

Rogowitz, B. E. & Voss, R. V. (1990). Shape perception and low-dimensional fractal boundary contours. *Human vision and electronic imaging: Models, methods, and applications, Proceedings SPIE, 1249*, 389–394.

Rovamo, J. & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research, 37*, 495–510.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22*, 325–345.

Strasburger, H., Harvey, L. O. Jr & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics, 49*, 495–508.

Strasburger, H., Rentschler, I. & Harvey, L. O. Jr (1994). Cortical magnification theory fails to predict visual recognition. *European Journal of Neuroscience, 6*, 1583–1588.

Takayasu, H. (1989). *Fractals in physical science*. Manchester: Manchester University Press.

Treutwein, B. & Rentschler, I. (1992). Double-pulse resolution in the visual field: The influence of temporal stimulus characteristics. *Clinical Vision Sciences, 7*, 421–434.

Van Essen, D. C., Newsome, W. T. & Maunsell, H. R. (1984). The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. *Vision Research, 24*, 429–448.

Virsu, V., Näsänen, R. & Osmoviita, K. (1987). Cortical magnification and peripheral vision. *Journal of the Optical Society of America A, 4*, 1568–1578.

Virsu, V. & Rovamo, J. (1979). Visual resolution, contrast sensitivity and the cortical magnification factor. *Experimental Brain Research, 37*, 475–494.

Virsu, V., Rovamo, J., Laurinen, P. & Näsänen, R. (1982). Temporal contrast sensitivity and cortical magnification. *Vision Research, 22*, 1211–1217.

Watanabe, S. (1985). *Pattern recognition. Human and mechanical*. New York: Wiley.

Wegmann, B. & Zetzsche, C. (1990). Visual system based polar quantization of local amplitude and local phase of orientation filter outputs. *Human vision and electronic imaging: Models, methods, and applications, Proceedings SPIE, 1249*, 607–613.

Wertheim, T. (1894). Über die indirekte Sehschärfe. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 7*, 172–187.

Westheimer, G. (1982). The spatial grain of the perifoveal visual field. *Vision Research, 22*, 157–162.

Weymouth, F. W. (1958). Visual sensory units and the minimal angle of resolution. *American Journal of Ophthalmology, 46*, 102–113.

Wilson, H. R., Levi, D., Maffei, L., Rovamo, J. & DeValois, R. (1990). The perception of form: Retina to striate cortex. In Spillmann, L. & Werner, J. B. (Eds), *Visual perception: The neurophysiological foundations* (Chap. 15). New York: Academic Press.

Zeevi, Y. Y. & Porat, M. (1989). Image representation by localized phase. *Visual communications and image processing, Proceedings SPIE, 1199*, 1512–1517.

Zetzsche, C. & Barth, E. (1990). Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Research, 30*, 1111–1117.

Acknowledgements—This study was supported by the Deutsche Forschungsgemeinschaft grants Re 337/5-3 and Re 337/10-1 to IR. The authors are grateful to Erhardt Barth, Mario Ferraro and Christoph Zetzsche for helpful comments on the manuscript.

APPENDIX I

Bayesian model for supervised learning and classification

Probabilistic virtual prototypes. Our classification model (see Rentschler *et al.*, 1994) is based on the parametric Bayesian approach of technical pattern recognition (e.g. Duda & Hart, 1973; Ahmed & Rao, 1975). It is assumed that the probability p_{ik} of a given signal vector $\mathbf{x}_i$ being assigned to class ω_k can be formally described as a posteriori classification probability for that class, i.e.

$$p_{ik} \equiv p(\omega_k|\mathbf{z}_i), \tag{A1}$$

where $i \in \{1,...,n_s\}$ and $k \in \{1,...,n_c\}$, with n_s being the total number of signals, n_c being the number of classes and $\mathbf{z}_i$ referring to the internal

representation of $\mathbf{x}_i$. This formulation implies that the characteristics of human classification behaviour do not directly depend on the physical input (i.e., the signal vector $\mathbf{x}$) but on its *perceptual realization* $\mathbf{z}$, which may be regarded as the result of an *internal measurement* process acting upon $\mathbf{x}$.

As to the relationships between physical and internal signal representations, we model them in terms of an additive error vector $\mathbf{w}$, i.e.

$$\mathbf{z} \equiv \mathbf{x} + \mathbf{w}. \tag{A2}$$

This approach is standard in applied optimal estimation (e.g. Gelb, 1974) and multidimensional extensions of signal detection theory (Ashby, 1992). It dates back to methodologies developed by Gauss (1809). The error vector $\mathbf{w}$ in A(2) incorporates the additional degrees of freedom of both *bias* and *variance* introduced by the perceptual process of internal feature measurement. Following the parametric approach, we assume this error signal to be normally distributed, i.e.

$$p(\mathbf{w}|\omega_j) = N(\mathbf{w}; \mu_j', \mathbf{C}_j'), \tag{A3}$$

where

$$N(\mathbf{w}; \mu_j', \mathbf{C}_j') = \frac{1}{\sqrt{|\mathbf{C}_j'|(2\pi)^{d/2}}} exp\left[-\frac{1}{2}(\mathbf{w} - \mu_j')^T \mathbf{C}_j'^{-1}(\mathbf{w} - \mu_j')\right]. \tag{A4}$$

Here μ_j' denotes the d-component expectation or mean vector of the jth class, $\mathbf{C}'_j$ is the corresponding $d \times d$ covariance matrix, $\mathbf{C}'^{-1}_j$ is the inverse of $\mathbf{C}_j$ and $|\mathbf{C}'_j|$ its determinant. The superscript T denotes transposition.

If we further assume $\mathbf{x}$ to be normally distributed with mean value μ_j and covariance $\mathbf{C}_j$ and provide statistical independence of $\mathbf{x}$ and $\mathbf{w}$, then we have as a resulting distribution of $\mathbf{z}$

$$p(\mathbf{z}|\omega_j) = N(\mathbf{z}; \mu_j + \mu_j', \mathbf{C}_j + \mathbf{C}_j'), \tag{A5}$$

which is also normal with the parameters $\mu_j + \mu_j'$ and $\mathbf{C}_j + \mathbf{C}_j'$. The assumption of normal or Gaussian distribution functions for internal signal representations belonging to a particular class is also motivated by reports that subjects enter categorization tasks with the expectation that the exemplars of each category are symmetrically and unimodally distributed around some prototype (Fried & Holyoak, 1984; Flannagan *et al.*, 1986).

According to Bayes theorem, $p(\omega_j|\mathbf{z})$ is given by

$$p(\omega_j|\mathbf{z}) = \frac{p(\mathbf{z}|\omega_j)P(\omega_j)}{\sum_k p(\mathbf{z}|\omega_k)P(\omega_k)}, \tag{A6}$$

where $P(\omega_j$ is the a priori *probability* of class ω_j and the summation in the denominator is over the number of classes.

If one assumes equal a priori probabilities for the different classes this expression is formally similar to the similarity-choice models of Shepard (1957) and Luce (1963). However, it explicitly states that the characteristics of human classification behaviour do not directly depend on physical input, but on its corresponding internal representation.

(A5) and (A6) describe a *PVP* model with four parameters: μ_j, μ_j', $\mathbf{C}_j$ and $\mathbf{C}_j'$. Parameters μ_j and $\mathbf{C}_j$ are completely specified by the distributions of the signals in physical feature space, whereas μ_j' and $\mathbf{C}_j'$ are free parameters of the model: μ_j' captures the additional degrees of *bias*, or distortion, and μ_j' that of *covariance*, or fuzziness, of the internal representation of class ω_j.

Optimization procedure. In principle, both the mean vectors μ_j' and the covariance matrix $\mathbf{C}_j'$ consisting of $d + d(d+1)/2$ components per class may be estimated by minimizing the mean-squared error between the predicted classification probabilities p'_{ik} [(A1) and (A6)] and the observed classification probabilities p_{ik}:

$$\sum_{i,k}(p_{ik} - p'_{ik})^2 = \min$$
$$\text{under the constraint } \sum_k p'_{ik} - 1 = 0, i\epsilon\{1, ..., n_s\}. \tag{A7}$$

Instead of this general bias/variance PVP-model, it is also possible to confine the variation to only one parameter, either μ_j' or $\mathbf{C}_j'$. This results in models restricted to the aspects of *bias* or *variance*, respectively. In the analysis of the experiments described in the following section we employed a *pure bias model*, i.e. we set

$$\mathbf{C}_j' = 0, j\epsilon\{1, ..., n_c\}.$$

This is equivalent to the assumption that the class conditional probability densities in physical and internal feature space have the same covariances and differ only in their mean values—a restriction that was due to the technical constraints of our equipment. We used random sequences of 15 signals at most in the experiments. This would seem to be a set of samples being too small for reasonably robust estimates of *both* μ_j' and $\mathbf{C}_j'$—a problem related to the so-called bias/variance dilemma of statistical inference (Geman *et al.*, 1992).

To solve the constrained optimization problem given by (A7) the components of μ_j' are estimated by computing the minimum of the mean-squared error,

$$e^2 = (n_c n_s - 1)_{-1} \sum_{i,j}(p'_{ij} - p_{ij})^2, \tag{A8}$$

via a direct search method [analysis program STEPIT (Chandler, 1969)].

Uniqueness of virtual prototype solutions. The predictions of the general *bias/variance* PVP-classifier are invariant against scale transformations, translations and rotations of the feature coordinate system. This is obvious from the fact that such transformations do not alter the *relative* sizes, positions and orientations (as determined by the class specific covariance matrices) of the class conditional densities [see (A5)].

Another issue is that of the uniqueness of the virtual prototype solutions as obtained with the *bias* model (see above). This model assumes that the covariance matrices are determined by the physical signal properties and, therefore, remain fixed. Consequently, any rotation or scaling transformation of a corresponding virtual prototype solution would alter the predicted classification probabilities. This is not true for translations because this type of transformation leaves the covariance matrices unchanged. It follows that the solutions of the bias model are unique with respect to orientation and scale [these two being determined by the μ_j and $\mathbf{C}_j$ parameters in (A5)] but not with respect to translations. This is to say, that the orientation and size of virtual prototype configurations relatively to the physical signal configurations are significant, whereas their relative position is not.

Learning tomograms. From a computational point of view, the analysis of human classification behaviour within the PVP concept is based on the confusion matrix representing the experimental classification data. This matrix can be defined in various ways: the *average* classification performance is derived from the cumulative classification matrix which combines the responses of the observers during the entire learning procedure. But averaging inevitably leads to a smoothing of the learning data thus obscuring rapid changes in classification behaviour. This might be avoided by determining a series of virtual prototypes on the sole basis of quasi instantaneous classification matrices, each of which is based on the responses of only a few consecutive trials. However, the fewer the number of observations (i.e., the more the estimates become localized in time) the more the classification matrices become influenced by statistical fluctuations which makes it more difficult to trace the systematic changes due to learning.

We chose to alleviate the problem of antagonistic effects of random fluctuations in behaviour and of learning by employing a technique of averaging within running, i.e., temporal Gaussian windows. Let R_i^j designate the subject's classification response to pattern i in the jth trial, R_i^j being either 1 in the case of a correct classification or 0 otherwise. To obtain an estimate of the classification probability P_i^j, i.e., the probability of a correct classification of pattern i at trial j, a weighted average of the responses R_i^j in the temporal neighbourhood of trial j is calculated. The weights are determined by a truncated Gaussian window function of width h. For the averaging it is therefore

sufficient to restrict the summation to an interval of width $2h$. Thus we have

$$P_i^j = \frac{1}{C}\sum_{k=j-h}^{j+h} R_i^k \exp -\frac{1}{2}\left(\frac{k-j}{h/2}\right)^2, \quad \text{(A9)}$$

where C is a normalization constant given by

$$C = \sum_{k=j-h}^{j+h} \exp -\frac{1}{2}\left(\frac{k-j}{h/2}\right)^2.$$

This means that the estimation of P_i^j at trial j primarily depends on the response on trial j. However, it also affected by both the preceding and the following responses within the given temporal window, although it still remains temporally localized at trial j due to the shape of the weighting function. The fact that both past and future responses contribute to the estimate in a symmetrical way provides continuity and coherence between consecutive estimates.

The series of virtual prototype configurations, which are based on running averages of the behavioral classification matrices, we call *learning tomograms*. They allow us to account for the dynamics of the learning process and can be used to trace the temporal evolution of the underlying internal representations.

APPENDIX II

The characterization of internal representations by their virtual perceptual dimension

The mathematical concept of *dimension* provides a natural means to make the notion of the degree of freedom more explicit. For instance, in a Euclidean three-dimensional space each point has 3 degrees of freedom, but if a set of points forms a surface all points must satisfy a constraint F(x,y,z) for some function F. This constraint reduces the degrees of freedom to 2, i.e., to the topological dimension of a surface. The necessity to extend the classical notion of the (integer-valued) Euclidean and topological dimension arises from the fact that there are more complex cases, e.g. the Peano curve or the triadic Koch curve, where the traditional concept runs into theoretical difficulties [for a more thorough discussion of these fractal sets and the related concept of self-similarity, see Mandelbrot (1982)].

A possible way to generalize the dimension to fractional values [in fact there exists quite a number of different approaches, cf. Takayasu (1989)], is the introduction of the capacity dimension, originally invented by Kolmogorov. Like the Hausdorff dimension it is based on the idea of coverings. Let the considered shape be a bounded set in a d-dimensional Euclidean space. Cover the set by d-dimensional spheres of identical radius $1/\varepsilon$. The capacity dimension D_C is given by

$$D_C = \lim_{\varepsilon \to 0} \frac{logN(\varepsilon)}{log1/\varepsilon} \quad \text{(B1)}$$

where $N(\varepsilon)$ denotes the minimal number of spheres.

From a mathematical point of view, D_C is well-defined now, but difficulties occur if this definition is applied to physical problems, because according to B(1) the dimension is defined only in the limit $\varepsilon \to 0$. A length of size 0, however, is an unphysical concept as a consequence of the uncertainty principle. Hence, there are experimental constraints on the observed scale range for which the definition is applied.

Many physically feasible methods of defining fractal dimensions have recently been devised (for a review see Feder, 1988; Takayasu, 1989). In the following we base the definition of the perceptual dimension on the so-called box-counting method. According to this method, we first divide the feature space into squares of edge length r, then count the number of cells $N(r)$ needed to cover the triangle shaped configurations of the virtual prototypes. This procedure is then repeated for a series of bisections of the side length of the squares. While the upper limit of the squares is given by the distance of two prototypes in feature space, the lower limit is given by the mean covariance of the sample distributions which may be regarded to present a measure of the coarse graining of the feature space. In the so-defined scale range, $N(r)$ satisfies the relation

$$N(r) \propto r^{-D\psi} \quad \text{(B2)}$$

when r is changed. D_ψ we call the *perceptual dimension* of the configuration. Note that because of (B1), D_ψ is equal to the slope in a logN(ε) vs log($1/\varepsilon$) plot. D_ψ may be considered as an index characterizing how far the configuration of the virtual prototypes employs the degrees of freedom of the feature space, given a scaling range confined by the upper boundary of the (macroscopic) between-class distance and the lower boundary of the (microscopic) within-class variance, where the corresponding dimension of the physical prototypes provides the standard of reference. The latter dimension, D_ϕ, is obtained from physical signal distribution in the same way as D_ψ is derived from the virtual prototype configuration.